

RAW SEQUENCE LISTING

**The Biotechnology Systems Branch of the Scientific and Technical
Information Center (STIC) no errors detected.**

Application Serial Number: 10/593,181
Source: IFWP
Date Processed by STIC: 10/05/2006

ENTERED



IFWP

RAW SEQUENCE LISTING

DATE: 10/05/2006

PATENT APPLICATION: US/10/593,181

TIME: 10:17:06

Input Set : E:\Final Sequence List-13987-00022-US.txt

Output Set: N:\CRF4\10052006\J593181.raw

3 <110> APPLICANT: Hillebrand, Helke
 4 Ebneeth, Marcus
 5 Nasholm, Torgny
 6 Erikson, Oskar
 7 Hertzberg, Magnus
 9 <120> TITLE OF INVENTION: IMPROVED CONSTRUCTS FOR MARKER EXCISION BASED ON DUAL-
 FUNCTION
 10 SELECTION MARKER
 12 <130> FILE REFERENCE: 13987-00022-US
 C--> 14 <140> CURRENT APPLICATION NUMBER: US/10/593,181
 C--> 14 <141> CURRENT FILING DATE: 2006-09-15
 14 <150> PRIOR APPLICATION NUMBER: PCT/EP2005/002734
 15 <151> PRIOR FILING DATE: 2005-03-15
 17 <150> PRIOR APPLICATION NUMBER: EP 04006358.8
 18 <151> PRIOR FILING DATE: 2004-03-17
 20 <160> NUMBER OF SEQ ID NOS: 78
 22 <170> SOFTWARE: PatentIn version 3.3
 24 <210> SEQ ID NO: 1
 25 <211> LENGTH: 1160
 26 <212> TYPE: DNA
 27 <213> ORGANISM: Rhodosporidium toruloides
 29 <220> FEATURE:
 30 <221> NAME/KEY: CDS
 31 <222> LOCATION: (1)..(1104)
 32 <223> OTHER INFORMATION: coding for DAAO
 34 <400> SEQUENCE: 1
 35 atg cac tcg cag aag cgc gtc gtt gtc ctc gga tca ggc gtt atc ggt 48
 36 Met His Ser Gln Lys Arg Val Val Val Leu Gly Ser Gly Val Ile Gly
 37 1 5 10 15
 39 ctg agc agc gcc ctc atc ctc gct cgg aag ggc tac agc gtg cat att 96
 40 Leu Ser Ser Ala Leu Ile Leu Ala Arg Lys Gly Tyr Ser Val His Ile
 41 20 25 30
 43 ctc gcg cgc gac ttg ccg gag gac gtc tcg agc cag act ttc gct tca 144
 44 Leu Ala Arg Asp Leu Pro Glu Asp Val Ser Ser Gln Thr Phe Ala Ser
 45 35 40 45
 47 cca tgg gct ggc gcg aat tgg acg cct ttc atg acg ctt aca gac ggt 192
 48 Pro Trp Ala Gly Ala Asn Trp Thr Pro Phe Met Thr Leu Thr Asp Gly
 49 50 55 60
 51 cct cga caa gca aaa tgg gaa gaa tcg act ttc aag aag tgg gtc gag 240
 52 Pro Arg Gln Ala Lys Trp Glu Glu Ser Thr Phe Lys Lys Trp Val Glu
 53 65 70 75 80
 55 ttg gtc ccg acg ggc cat gcc atg tgg ctc aag ggg acg agg cgg ttc 288
 56 Leu Val Pro Thr Gly His Ala Met Trp Leu Lys Gly Thr Arg Arg Phe
 57 85 90 95

RAW SEQUENCE LISTING

DATE: 10/05/2006

PATENT APPLICATION: US/10/593,181

TIME: 10:17:06

Input Set : E:\Final Sequence List-13987-00022-US.txt

Output Set: N:\CRF4\10052006\J593181.raw

```

59 gcg cag aac gaa gac ggc ttg ctc ggg cac tgg tac aag gac atc acg      336
60 Ala Gln Asn Glu Asp Gly Leu Leu Gly His Trp Tyr Lys Asp Ile Thr
61      100      105      110
63 cca aat tac cgc ccc ctc cca tct tcc gaa tgt cca cct ggc gct atc      384
64 Pro Asn Tyr Arg Pro Leu Pro Ser Ser Glu Cys Pro Pro Gly Ala Ile
65      115      120      125
67 ggc gta acc tac gac acc ctc tcc gtc cac gca cca aag tac tgc cag      432
68 Gly Val Thr Tyr Asp Thr Leu Ser Val His Ala Pro Lys Tyr Cys Gln
69      130      135      140
71 tac ctt gca aga gag ctg cag aag ctc ggc gcg acg ttt gag aga cgg      480
72 Tyr Leu Ala Arg Glu Leu Gln Lys Leu Gly Ala Thr Phe Glu Arg Arg
73 145      150      155      160
75 acc gtt acg tcg ctt gag cag gcg ttc gac ggt gcg gat ttg gtg gtc      528
76 Thr Val Thr Ser Leu Glu Gln Ala Phe Asp Gly Ala Asp Leu Val Val
77      165      170      175
79 aac gct acg gga ctt ggc gcc aag tcg att gcg ggc atc gac gac caa      576
80 Asn Ala Thr Gly Leu Gly Ala Lys Ser Ile Ala Gly Ile Asp Asp Gln
81      180      185      190
83 gcc gcc gag cca atc cgc ggg caa acc gtc ctc gtc aag tcc cca tgg      624
84 Ala Ala Glu Pro Ile Arg Gly Gln Thr Val Leu Val Lys Ser Pro Cys
85      195      200      205
87 aag cga tgc acg atg gac tcg tcc gac ccc gct tct ccc gcc tac atc      672
88 Lys Arg Cys Thr Met Asp Ser Ser Asp Pro Ala Ser Pro Ala Tyr Ile
89      210      215      220
91 att ccc cga cca ggt ggc gaa gtc atc tgc ggc ggg acg tac ggc gtg      720
92 Ile Pro Arg Pro Gly Gly Glu Val Ile Cys Gly Gly Thr Tyr Gly Val
93 225      230      235      240
95 gga gac tgg gac ttg tct gtc aac cca gag acg gtc cag cgg atc ctc      768
96 Gly Asp Trp Asp Leu Ser Val Asn Pro Glu Thr Val Gln Arg Ile Leu
97      245      250      255
99 aag cac tgc ttg cgc ctc gac ccg acc atc tcg agc gac gga acg atc      816
100 Lys His Cys Leu Arg Leu Asp Pro Thr Ile Ser Ser Asp Gly Thr Ile
101      260      265      270
103 gaa ggc atc gag gtc ctc cgc cac aac gtc ggc ttg cga cct gca cga      864
104 Glu Gly Ile Glu Val Leu Arg His Asn Val Gly Leu Arg Pro Ala Arg
105      275      280      285
107 cga ggc gga ccc cgc gtt gag gca gaa cgg atc gtc ctg cct ctc gac      912
108 Arg Gly Gly Pro Arg Val Glu Ala Glu Arg Ile Val Leu Pro Leu Asp
109      290      295      300
111 cgg aca aag tcg ccc ctc tcg ctc ggc agg ggc agc gca cga gcg gcg      960
112 Arg Thr Lys Ser Pro Leu Ser Leu Gly Arg Gly Ser Ala Arg Ala Ala
113 305      310      315      320
115 aag gag aag gag gtc acg ctt gtg cat gcg tat ggc ttc tcg agt gcg      1008
116 Lys Glu Lys Glu Val Thr Leu Val His Ala Tyr Gly Phe Ser Ser Ala
117      325      330      335
119 gga tac cag cag agt tgg ggc gcg gcg gag gat gtc gcg cag ctc gtc      1056
120 Gly Tyr Gln Gln Ser Trp Gly Ala Ala Glu Asp Val Ala Gln Leu Val
121      340      345      350
123 gac gag gcg ttc cag cgg tac cac ggc gcg gcg cgg gag tcg aag ttg      1104

```

RAW SEQUENCE LISTING

DATE: 10/05/2006

PATENT APPLICATION: US/10/593,181

TIME: 10:17:06

Input Set : E:\Final Sequence List-13987-00022-US.txt

Output Set: N:\CRF4\10052006\J593181.raw

```

124 Asp Glu Ala Phe Gln Arg Tyr His Gly Ala Ala Arg Glu Ser Lys Leu
125          355          360          365
127 tagggcgggg tttgtggctg tattgcgggc atctacaaga aaaaaaaaaa aaaaaa      1160
130 <210> SEQ ID NO: 2
131 <211> LENGTH: 368
132 <212> TYPE: PRT
133 <213> ORGANISM: Rhodosporidium toruloides
135 <400> SEQUENCE: 2
136 Met His Ser Gln Lys Arg Val Val Val Leu Gly Ser Gly Val Ile Gly
137 1          5          10          15
139 Leu Ser Ser Ala Leu Ile Leu Ala Arg Lys Gly Tyr Ser Val His Ile
140          20          25          30
142 Leu Ala Arg Asp Leu Pro Glu Asp Val Ser Ser Gln Thr Phe Ala Ser
143          35          40          45
145 Pro Trp Ala Gly Ala Asn Trp Thr Pro Phe Met Thr Leu Thr Asp Gly
146          50          55          60
148 Pro Arg Gln Ala Lys Trp Glu Glu Ser Thr Phe Lys Lys Trp Val Glu
149 65          70          75          80
151 Leu Val Pro Thr Gly His Ala Met Trp Leu Lys Gly Thr Arg Arg Phe
152          85          90          95
154 Ala Gln Asn Glu Asp Gly Leu Leu Gly His Trp Tyr Lys Asp Ile Thr
155          100          105          110
157 Pro Asn Tyr Arg Pro Leu Pro Ser Ser Glu Cys Pro Pro Gly Ala Ile
158          115          120          125
160 Gly Val Thr Tyr Asp Thr Leu Ser Val His Ala Pro Lys Tyr Cys Gln
161          130          135          140
163 Tyr Leu Ala Arg Glu Leu Gln Lys Leu Gly Ala Thr Phe Glu Arg Arg
164 145          150          155          160
166 Thr Val Thr Ser Leu Glu Gln Ala Phe Asp Gly Ala Asp Leu Val Val
167          165          170          175
169 Asn Ala Thr Gly Leu Gly Ala Lys Ser Ile Ala Gly Ile Asp Asp Gln
170          180          185          190
172 Ala Ala Glu Pro Ile Arg Gly Gln Thr Val Leu Val Lys Ser Pro Cys
173          195          200          205
175 Lys Arg Cys Thr Met Asp Ser Ser Asp Pro Ala Ser Pro Ala Tyr Ile
176          210          215          220
178 Ile Pro Arg Pro Gly Gly Glu Val Ile Cys Gly Gly Thr Tyr Gly Val
179 225          230          235          240
181 Gly Asp Trp Asp Leu Ser Val Asn Pro Glu Thr Val Gln Arg Ile Leu
182          245          250          255
184 Lys His Cys Leu Arg Leu Asp Pro Thr Ile Ser Ser Asp Gly Thr Ile
185          260          265          270
187 Glu Gly Ile Glu Val Leu Arg His Asn Val Gly Leu Arg Pro Ala Arg
188          275          280          285
190 Arg Gly Gly Pro Arg Val Glu Ala Glu Arg Ile Val Leu Pro Leu Asp
191          290          295          300
193 Arg Thr Lys Ser Pro Leu Ser Leu Gly Arg Gly Ser Ala Arg Ala Ala
194 305          310          315          320
196 Lys Glu Lys Glu Val Thr Leu Val His Ala Tyr Gly Phe Ser Ser Ala

```

RAW SEQUENCE LISTING

DATE: 10/05/2006

PATENT APPLICATION: US/10/593,181

TIME: 10:17:06

Input Set : E:\Final Sequence List-13987-00022-US.txt

Output Set: N:\CRF4\10052006\J593181.raw

```

197          325          330          335
199 Gly Tyr Gln Gln Ser Trp Gly Ala Ala Glu Asp Val Ala Gln Leu Val
200          340          345          350
202 Asp Glu Ala Phe Gln Arg Tyr His Gly Ala Ala Arg Glu Ser Lys Leu
203          355          360          365
205 <210> SEQ ID NO: 3
206 <211> LENGTH: 1005
207 <212> TYPE: DNA
208 <213> ORGANISM: Caenorhabditis elegans
210 <220> FEATURE:
211 <221> NAME/KEY: CDS
212 <222> LOCATION: (1)..(1002)
213 <223> OTHER INFORMATION: coding for DAO
215 <400> SEQUENCE: 3
216 atg gca aac ata att ccg aag att gca att atc ggc gaa gga gtc att      48
217 Met Ala Asn Ile Ile Pro Lys Ile Ala Ile Ile Gly Glu Gly Val Ile
218 1          5          10          15
220 gga tgt act tca gca ctt caa ata tca aaa gct ata cca aat gcg aaa      96
221 Gly Cys Thr Ser Ala Leu Gln Ile Ser Lys Ala Ile Pro Asp Ala Lys
222          20          25          30
224 ata act gtg ctc cac gat aaa cca ttt aaa aaa tcg tgc agt gca gga      144
225 Ile Thr Val Leu His Asp Lys Pro Phe Lys Lys Ser Cys Ser Ala Gly
227          35          40          45
229 cca gca gga tta ttt aga atc gat tat gag gag aat act gaa tac gga      192
230 Pro Ala Gly Leu Phe Arg Ile Asp Tyr Glu Glu Asn Thr Glu Tyr Gly
231          50          55          60
233 cgt gct tct ttc gcc tgg ttc tca cat ctc tat cgc act aca aaa gga      240
234 Arg Ala Ser Phe Ala Trp Phe Ser His Leu Tyr Arg Thr Thr Lys Gly
235 65          70          75          80
237 tcc gaa acc ggc gtg aaa tta gtt tct gga cat att caa tcc gac aac      288
238 Ser Glu Thr Gly Val Lys Leu Val Ser Gly His Ile Gln Ser Asp Asn
239          85          90          95
241 ttg gag tca ttg aag caa caa caa aga gcc tat ggc gat att gtg tac      336
242 Leu Glu Ser Leu Lys Gln Gln Gln Arg Ala Tyr Gly Asp Ile Val Tyr
243          100          105          110
245 aac ttt aga ttc ttg gat gat aga gaa cgg ctg gac att ttt ccc gaa      384
246 Asn Phe Arg Phe Leu Asp Asp Arg Glu Arg Leu Asp Ile Phe Pro Glu
247          115          120          125
249 cca tca aag cac tgc att cac tac acc gcc tac gca tca gaa ggt aac      432
250 Pro Ser Lys His Cys Ile His Tyr Thr Ala Tyr Ala Ser Glu Gly Asn
251          130          135          140
253 aag tac gtg cct tat ttg aag aat ttg ctg ctt gag caa aaa atc gag      480
254 Lys Tyr Val Pro Tyr Leu Lys Asn Leu Leu Leu Glu Gln Lys Ile Glu
255 145          150          155          160
257 ttc aag caa caa gaa gtg acg agt ttg gac gca gtc gcc gac gct ggt      528
258 Phe Lys Gln Gln Glu Val Thr Ser Leu Asp Ala Val Ala Asp Ala Gly
260          165          170          175
262 tac gat gtt att gta aac tgc gca ggc ttg tac ggt gga aag ttg gct      576
263 Tyr Asp Val Ile Val Asn Cys Ala Gly Leu Tyr Gly Gly Lys Leu Ala

```

RAW SEQUENCE LISTING

DATE: 10/05/2006

PATENT APPLICATION: US/10/593,181

TIME: 10:17:06

Input Set : E:\Final Sequence List-13987-00022-US.txt

Output Set: N:\CRF4\10052006\J593181.raw

```

264          180          185          190
266 ggt gat gac gat act tgc tac ccc att aga gga gtc att ttg gaa gtt      624
267 Gly Asp Asp Asp Thr Cys Tyr Pro Ile Arg Gly Val Ile Leu Glu Val
268          195          200          205
270 gat gca cca tgg cac aag cac ttc aat tat cga gac ttt act act ttc      672
271 Asp Ala Pro Trp His Lys His Phe Asn Tyr Arg Asp Phe Thr Thr Phe
272          210          215          220
274 aca att cca aaa gag cac agc gtg gtg gtt ggg tcc acc aag cag gac      720
275 Thr Ile Pro Lys Glu His Ser Val Val Val Gly Ser Thr Lys Gln Asp
276 225          230          235          240
278 aat cga tgg gat ttg gag atc acc gac gag gat aga aat gat att ttg      768
279 Asn Arg Trp Asp Leu Glu Ile Thr Asp Glu Asp Arg Asn Asp Ile Leu
280          245          250          255
282 aaa cga tac att gct tta cat cct gga atg aga gag cca aag att atc      816
283 Lys Arg Tyr Ile Ala Leu His Pro Gly Met Arg Glu Pro Lys Ile Ile
284          260          265          270
286 aaa gaa tgg tca gca ctt cgc ccg gga cgt aag cat gtc aga att gaa      864
287 Lys Glu Trp Ser Ala Leu Arg Pro Gly Arg Lys His Val Arg Ile Glu
288          275          280          285
290 gcg cag aag agg aca tct gtt gga aac tca aaa gat tat atg gtt gtg      912
291 Ala Gln Lys Arg Thr Ser Val Gly Asn Ser Lys Asp Tyr Met Val Val
293          290          295          300
295 cat cac tat ggt cac ggg agc aac gga ttc acg ttg ggt tgg gga aca      960
296 His His Tyr Gly His Gly Ser Asn Gly Phe Thr Leu Gly Trp Gly Thr
297 305          310          315          320
299 gca att gaa gca act aaa ctt gtt aag act gca cta gga tta taa      1005
300 Ala Ile Glu Ala Thr Lys Leu Val Lys Thr Ala Leu Gly Leu
301          325          330
303 <210> SEQ ID NO: 4
304 <211> LENGTH: 334
305 <212> TYPE: PRT
306 <213> ORGANISM: Caenorhabditis elegans
308 <400> SEQUENCE: 4
309 Met Ala Asn Ile Ile Pro Lys Ile Ala Ile Ile Gly Glu Gly Val Ile
310 1          5          10          15
312 Gly Cys Thr Ser Ala Leu Gln Ile Ser Lys Ala Ile Pro Asn Ala Lys
313          20          25          30
315 Ile Thr Val Leu His Asp Lys Pro Phe Lys Lys Ser Cys Ser Ala Gly
316          35          40          45
318 Pro Ala Gly Leu Phe Arg Ile Asp Tyr Glu Glu Asn Thr Glu Tyr Gly
319          50          55          60
321 Arg Ala Ser Phe Ala Trp Phe Ser His Leu Tyr Arg Thr Thr Lys Gly
322 65          70          75          80
324 Ser Glu Thr Gly Val Lys Leu Val Ser Gly His Ile Gln Ser Asp Asn
325          85          90          95
327 Leu Glu Ser Leu Lys Gln Gln Gln Arg Ala Tyr Gly Asp Ile Val Tyr
328          100          105          110
330 Asn Phe Arg Phe Leu Asp Asp Arg Glu Arg Leu Asp Ile Phe Pro Glu
331          115          120          125

```

RAW SEQUENCE LISTING ERROR SUMMARY

DATE: 10/05/2006

PATENT APPLICATION: US/10/593,181

TIME: 10:17:07

Input Set : E:\Final Sequence List-13987-00022-US.txt

Output Set: N:\CRF4\10052006\J593181.raw

Please Note:

Use of n and/or Xaa have been detected in the Sequence Listing. Please review the Sequence Listing to ensure that a corresponding explanation is presented in the <220> to <223> fields of each sequence which presents at least one n or Xaa.

Seq#:17; Xaa Pos. 1,2,4,7,8,9,10,12,13,14,15,16,18

Seq#:22; N Pos. 25,26,27,28,29,30,31,32,33,34,35,36,37,38,39

Seq#:45; N Pos. 25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44

Seq#:45; N Pos. 45,46,47,48,49,50,51,52,53,54,55,56,57,58,59,60,61,62,63

Seq#:62; N Pos. 6,12

Seq#:66; N Pos. 25,26,27,28,29,30,31,32,33,34,35,36,37,38

VERIFICATION SUMMARY

DATE: 10/05/2006

PATENT APPLICATION: US/10/593,181

TIME: 10:17:07

Input Set : E:\Final Sequence List-13987-00022-US.txt

Output Set: N:\CRF4\10052006\J593181.raw

L:14 M:270 C: Current Application Number differs, Replaced Current Application No
L:14 M:271 C: Current Filing Date differs, Replaced Current Filing Date
L:1848 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:17 after pos.:0
L:1851 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:17 after pos.:16
L:1918 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:22 after pos.:0
L:2202 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:45 after pos.:0
L:2204 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:45 after pos.:60
L:2419 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:62 after pos.:0
L:2472 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:66 after pos.:0